High Impulse Noise Intensity Removal in Natural Images Using Convolutional Neural Network

Mehdi Mafi
Center for Advanced Technology and Education, Department of Electrical and Computer Engineering
Florida International University
Miami, FL, USA
mmafi002@fiu.edu

Walter Izquierdo
Center for Advanced Technology and Education, Department of Electrical and Computer Engineering
Florida International University
Miami, FL, USA
wizqu003@fiu.edu

Malek Adjouadi
Center for Advanced Technology and Education, Department of Electrical and Computer Engineering
Florida International University
Miami, FL, USA
adjouadi@fiu.edu

***Abstract*— This paper introduces a new image smoothing filter based on a feed-forward convolutional neural network (CNN) in presence of impulse noise. This smoothing filter integrates a very deep architecture, a regularization method, and a batch normalization process. This fully integrated approach yields an effectively denoised and smoothed image yielding a high similarity measure with the original noise free image. Specific structural metrics are used to assess the denoising process and how effective was the removal of the impulse noise. This CNN model can also deal with other noise levels not seen during the training phase. The proposed CNN model is constructed through a 20-layer network using 400 images from the Berkeley Segmentation Dataset (BSD) in the training phase. Results are obtained using the standard testing set of 8 natural images not seen in the training phase. The merits of this proposed method are weighed in terms of high similarity measure and structural metrics that conform to the original image and compare favorably to the different results obtained using state-of-art denoising filters.**

Keywords*— *Impulse noise, Convolutional Neural Network (CNN), Image, denoising

I. Introduction

Impulse noise is inherent to digital images and degrades the quality of images [1]. In this study, removal of impulse (salt and pepper) random noise is considered. It appears as white and black pixels in the noisy image. Assuming normalization [2][3][4], the salt and pepper noise model can be expressed by as follows:

$$I_c = \begin{cases} 0 & Probability\ P_p \\ 1 & Probability\ P_s \\ C & Probability\ 1 - P_p - P_s \end{cases} \quad (1)$$

Where C is uncorrupted pixels, probabilities P_s (Salt) and P_p (Pepper) are assigned to corrupted pixels.

When removing noise in images, one of the most challenging part is in smoothing the image to attenuate the noise effect while preserving the image details. In most instances, smoothing filters cause irreversible loss of important details, affecting as a consequence edges and object boundaries. Therefore, an effectual smoothing filter is one that minimizes the effect of noise no matter its intensity while at the same time it preserves the image details. The effectiveness of smoothing filter is measured through high similarity and high structural metrics to the clean image. The proposed smoothing filter is based on a deep convolutional neural network (CNN) designed to remove the impulse (salt and pepper) noise and the latent original image is directly estimated. Then, batch normalization is applied in order to speed up and improve the denoising process. Finally, the network is trained for unknown (blind) noise.

The results are compared to another well-established denoising filter, the adaptive median and fixed weighted mean filter AMFWMF [3][4], which was already proven to outperform many of the relevant denoising methods introduced in the literature [5][6][7][8][9][10][11][12][13]. Several structural metrics, including peak signal to noise ratio (PSNR), structural similarity index (SSIM) and feature similarity index (FSIM) are used for evaluation purposes, and to ensure that any ensuing edge detection method that follows such denoising process preserves edge details. The aim of this study is to determine if the PSNR, SSIM and FSIM metrics can be improved through the use of convolutional neural networks over the use of an image processing based algorithm we named the switching adaptive median and fixed weighted mean filter introduced in [3].

II. Proposed Method

In this section, the network architecture and evaluation measures are discussed. The *network architecture* subsection provides the design details of the CNN architecture developed to address the effects of impulse noise in digital images. The *evaluation measures* subsection provides the mathematical foundation for the structural metrics and the peak signal to noise ratio as used in this study for evaluation purposes. The genesis for this CNN design is in discovering that most existing filters don't have a good performance at the boundaries, especially in the presence of high noise level, in spite of having good performance overall in terms of noise removal. So the proposed design aims at preserving all image details including boundaries and yielding at the same time high PSNR and structural similarities.

A. Network Architecture

The proposed CNN model is a modification on the DnCNN introduced in [14] for impulse denoising. Then, training data using 400 images from the well-known *Berkeley Segmentation Dataset* (BSD) is provided for learning. The input to the CNN is a noisy image y_i produced by artificially injecting impulse noise over the clean original image (x_i) and the output is directly estimated as the denoised original image as $f(y_i)$. The l-2 loss function is the summation of squared error between the estimated latent noise-free original patches (mapped noisy patches) and the noise-free original patches as

National Science Foundation (NSF)
Dissertation Year Fellowship (DYF)
978-1-7281-3783-4/20/$31.00 ©2020 IEEE

given in (2), then, the network parameter are updated by minimizing the loss function [14][15].

$$L = \sum_{i=1}^{N} ||f(y_i) - x_i||_2^2 \quad (2)$$

Where N is number of training patches set ($\{y_i, x_i\}$). The network model introduced in this study follows the procedures in [14] and has 3 types of layers: 1st layer is a combination of convolution and ReLU (for non-linearity) [16] in which 64 filters of size 3×3×1) are used in order to create 64 feature maps. 2nd layer to $(Depth\ of\ the\ network - 1)^{th}$ layer is a combination of convolutional, batch normalization [17] and rectified linear unit (ReLU) [16] in which 64 filters (size of 3×3×64) are used. For reconstruction purposes, one filter of size 3×3×64 is used at the last layer as a convolutional layer. Combination of convolution and ReLU [16] separates the noise from noisy observations through the hidden layers. Also, it is noted that size mismatches between different input images can cause boundary artifacts [14]. Therefore, the input image is directly padded with zeros before the first convolution stage in order to reduce such boundary artifacts. The network model is illustrated in Fig. 1

B. Evaluation Measures

The feature similarity index (FSIM) [18] between noisy and denoised images is measured as follows:

$$\text{FSIM} = \frac{\sum_{i\epsilon\Omega} S_L(i).PC_m(i)}{\sum_{i\epsilon\Omega} PC_m(i)} \quad (3)$$

Where Ω is the whole image spatial domain, $PC_m(i) = \max(PC_1(i), PC_2(i))$, $S_L(i)$ is the similarity at location i as (4):

$$S_L(i) = S_{PC}(i) S_G(i) \quad (4)$$

$S_{PC}(i)$ is the similarity measure between two feature maps (PC_1 and PC_2, they are extracted from noisy image and denoised image) as (5).

$$S_{PC}(i) = \frac{2PC_1(i).PC_2(i)+T_1}{PC_1(i)^2+PC_2(i)^2+T_1} \quad (5)$$

PC is the model of how HVS detect and identify features which is between 0 and 1, T_1 is a positive constant used in order to increase the stability of $S_{PC}(i)$.
$S_G(i)$ is the similarity measure between G_1 and G_2 (partial derivatives of image) as (6).

$$S_G(i) = \frac{2G_1(i).G_2(i)+T_2}{G_1(i)^2+G_2(i)^2+T_2} \quad (6)$$

T_2 is a positive constant, which depends on the dynamic range of $GM = \sqrt{{G_1}^2 + {G_2}^2}$.

The structural similarity index (SSIM) as defined in [19] is expressed as follows:

$$\text{SSIM} = \frac{(2\bar{x}\bar{y}+C1)(2\sigma_{xy}+C2)}{(\bar{x}^2+\bar{y}^2+C1)({\sigma_x}^2+{\sigma_y}^2+C2)} \quad (7)$$

Where σ_x and σ_y denote the standard deviations for the x and y images, respectively with the standard deviation of their combination expressed as σ_{xy}

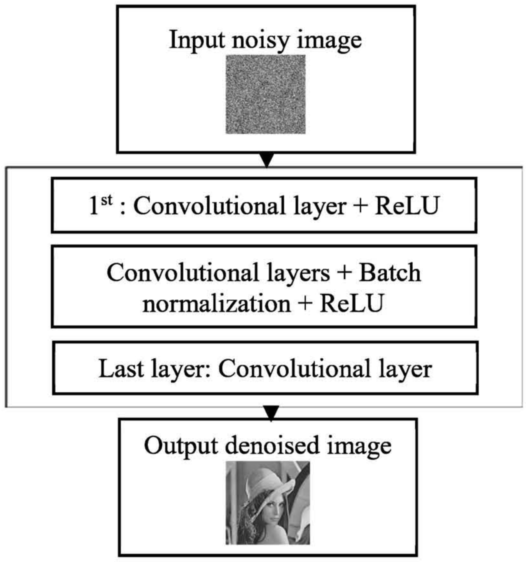


Fig. 1. Network model

The peak signal to noise ratio (PSNR) is defined as follows:

$$PSNR = 10log\frac{(\max(x))^2}{MSE} \quad (8)$$

Where MSE defines the mean square error in which max (x) denotes the maximum intensity of the pixels.

III. Result And Discussion

With the intent to preserve as much image details as possible while the effects of impulse noise are minimized, the results of the CNN-based denoising filter are compared to the results obtained with the most recent and highly effective AMFWMF detailed in [3]. The 400 (180×180) images from the Berkeley Segmentation Dataset (BSD) [14] are used in the training phase. Additionally, we set aside another 8 natural images as shown in Fig. 2 in the testing phase. It is important to emphasize that the images that are used for testing are not seen in the training phase.

As indicated earlier, the optimal results were obtained by using depth of 20 layers with 40×40 patch size for unknown denoising. The stochastic gradient descent (SGD)-momentum with weight decay of 0.0001, momentum of 0.9, and a mini-batch of 128 is used in similar fashion as in [14] [20] [21]. There are 50 epochs used for our model. MatConvNet package which is a MATLAB toolbox for Convolutional network (CNN) is used in this study. The implementation is carried out in MATLAB 2017b on a PC with Nvidia GPU. The training time of the network consisted nearly of 24 hours.

Tables I shows the results obtained on the averaged peak signal to noise ratio (PSNR), averaged structural similarity index (SSIM) and averaged FSIM measures, comparing the AMFWMF filter (the adaptive median filter of AMFWMF is set based on the minimum and maximum initial window size for the assumed noise level) against the proposed CNN-based method. All these metrics are calculated in the presence of different impulse noise, which varied in this case from 10 to 90 percent on the 8 aforementioned testing images.

Fig. 3 shows the same comparison in the presence of 90% impulse noise on test image "Lena". Fig. 4-6 show the proposed filter (unknown denoising) results in the presence of different impulse noise on different testing images. As the test images show, the proposed CNN-based filter yields very good results under different noise level intensities in terms of

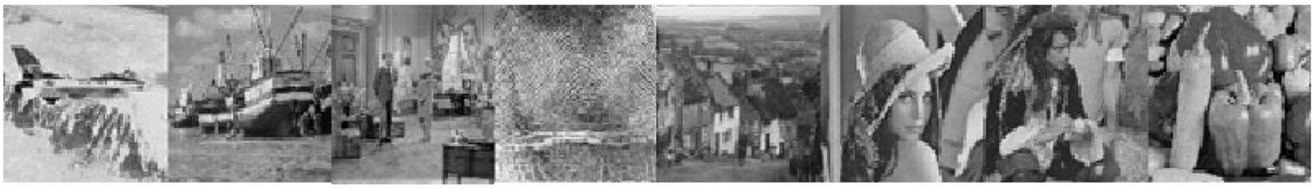


Fig. 2. Additional 8 test images

denoising, preservation of details and high similarity and quality of the denoised image with respect to the original noise-free image.

IV. SUMMARY

A new CNN-based denoising filter is proposed using a deep feed-forward convolutional neural network (CNN) filter with an appropriate loss function. The focus is placed on removing or minimizing the impulse (salt and pepper) noise even at high intensity levels. This filter is consequently shown to yield the highest evaluation measures in contrast to the AMFWMF method [3][4] and the other well-known denoising filters that were compared earlier to the AMFWMF. The results show that the CNN method preserves more of the edge details as reflected by the highest structural similarity measures and the PSNR, proving the similitude of the denoised image to the clean image. This similitude under the CNN method seems to be resilient to the effects of impulse noise even under high intensity levels.

ACKNOWLEDGMENT

We are grateful for the continued support from the National Science Foundation (NSF) under grants CNS-1920182, CNS-1532061, CNS 1551221, and CNS 1338922. We remain grateful for the support provided through the Ware Foundation. Additional support is provided through the FIU-University Graduate School (UGS) through the dissertation year fellowship (DYF) provided to Mr. Mehdi Mafi.

REFERENCES

[1] M. Mafi, H. Martin, M. Cabrerizo, J. Andrian, A. Barreto, and M. Adjouadi, "A comprehensive survey on Impulse and Gaussian Denoising Filters for Digital Images", Signal Processing, vol. 157, pp. 236–260, 2019.

[2] K. M. Moon and M. D. Patil, B. Parmar, "Image Restoration Using Adaptive Switching Median Filter," In Proc. IEEE Conference on Computational Intelligence and Computing Research, pp. 1-4, 2010.

[3] M. Mafi, H. Rajaei, M. Cabrerizo and M. Adjouadi, "A Robust Edge Detection Approach in the Presence of High Impulse Noise Intensity through Switching Adaptive Median and Fixed Weighted Mean Filtering", IEEE Trans. Image Process., vol. 27, no. 11, pp. 5475 – 5490, Nov. 2018.

[4] M. Mafi, H. Martin, M. Adjouadi, "High Impulse Noise Intensity Removal in MRI Images," In Proc. IEEE Signal Processing in Medicine and Biology Symposium, Philadelphia, PA, Dec. 2017.

[5] I. F. Jafar, R. A. AlNa'mneh, K. A. Darabkh, "Efficient Improvements on the BDND Filtering Algorithm for the Removal of High-Density Impulse Noise," IEEE Trans. Image Process., vol. 22, no. 3, Mar. 2013.

TABLE I. AVERAGE PSNR, SSIM AND FSIM COMPARISON ON 8 TEST IMAGES

	AMFWMF			***Proposed CNN***		
	PSNR	*SSIM*	*FSIM*	***PSNR***	***SSIM***	***FSIM***
10%	28.8309	0.9698	0.9979	40.7748 ± 0.0125	$0.9869 \pm 4.4721 \times 10^{-5}$	$0.9980 \pm 4.5241 \times 10^{-5}$
20%	28.6718	0.9549	0.9965	37.9328 ± 0.0171	$0.9780 \pm 7.6089 \times 10^{-5}$	$0.9969 \pm 2.2361 \times 10^{-5}$
30%	26.3245	0.9437	0.9932	37.4581 ± 0.0148	$0.9684 \pm 7.9472 \times 10^{-5}$	$0.9955 \pm 5.1042 \times 10^{-5}$
40%	23.9689	0.9052	0.9926	36.1030 ± 0.0187	$0.9574 \pm 9.2338 \times 10^{-5}$	0.9939 ± 0.0013
50%	23.7025	0.8815	0.9902	34.7381 ± 0.0229	0.9466 ± 0.0112	$0.9909 \pm 5.1042 \times 10^{-5}$
60%	23.9220	0.8442	0.9838	33.2798 ± 0.0204	$0.9269 \pm 1.8778 \times 10^{-4}$	$0.9865 \pm 8.3351 \times 10^{-5}$
70%	23.2654	0.8331	0.9721	31.5983 ± 0.0335	$0.9027 \pm 2.4623 \times 10^{-4}$	$0.9790 \pm 1.4749 \times 10^{-4}$
80%	21.5890	0.7865	0.9521	29.5832 ± 0.0227	$0.8641 \pm 3.6746 \times 10^{-4}$	$0.9644 \pm 2.8266 \times 10^{-4}$
90%	20.0002	0.7201	0.9067	26.6888 ± 0.0346	$0.7835 \pm 5.1186 \times 10^{-4}$	$0.9269 \pm 5.1186 \times 10^{-4}$

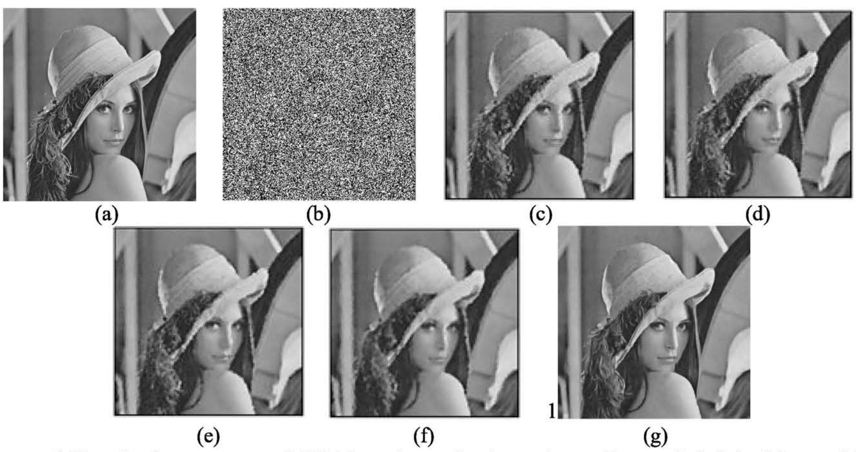
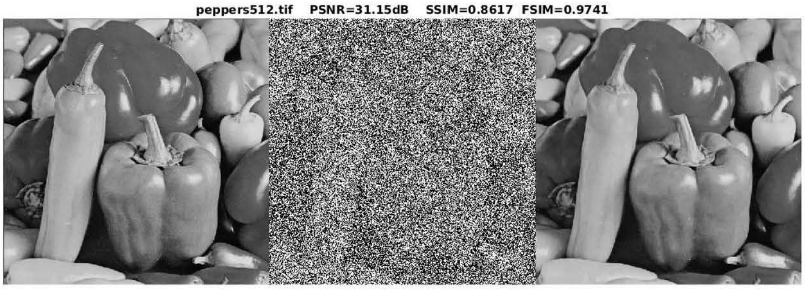
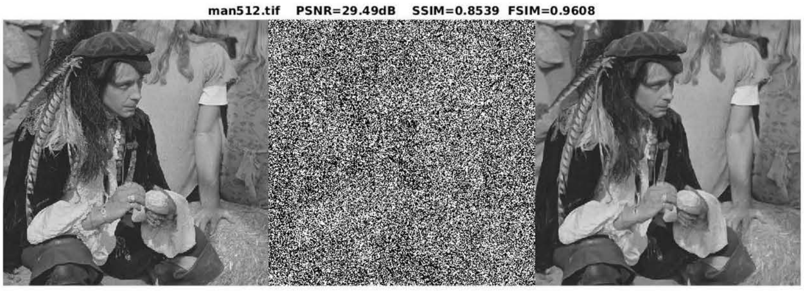


Fig. 3. Applying the proposed filter in the presence of 90% impulse noise intensity on Lena, a) Original image b) Noisy image c) AMFWMF (initial window size=3) d) AMFWMF (initial adaptive median window size=5) e)) AMFWMF (initial adaptive median window size=7) f)) AMFWMF (initial adaptive median window size=9) g) Proposed filter (unknown denoising)

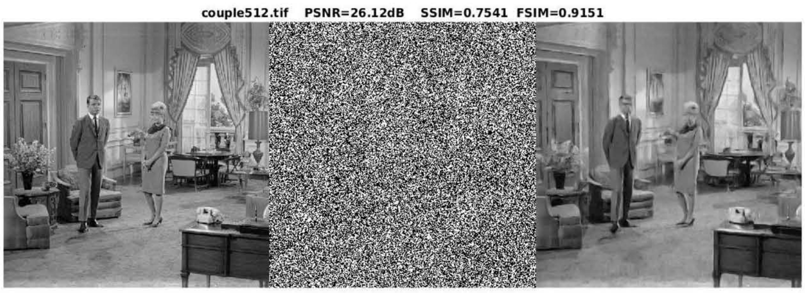


(a)

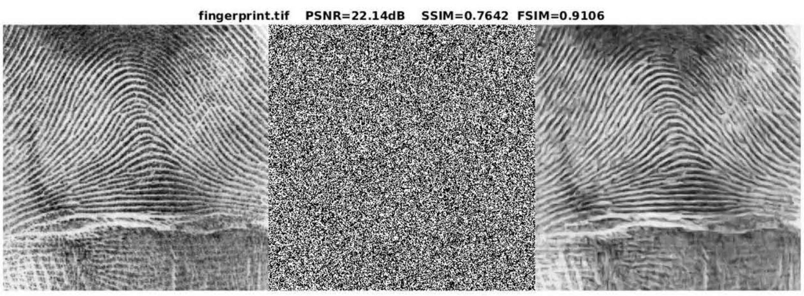


(b)

Fig. 4. Results of proposed filter (unknown denoising) on different test images, columns 1 through 3 are: Original test image, noisy image, and denoised image a) Test image “Fruits” corrupted with 80% impulse noise b) Test image “Man” corrupted with 80% impulse noise

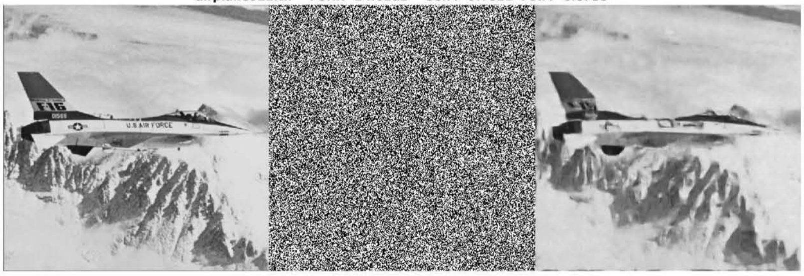


(a)

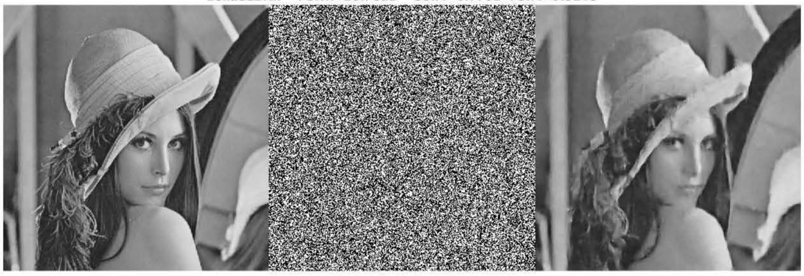


(b)

Fig. 5. Results of the proposed filter (unknown denoising) on different test images, columns 1 through 3 are: Original test image, noisy image, and denoised image a) Test image “Couple” corrupted with 90% impulse noise b) Test image “Finger print” corrupted with 90% impulse noise



(a)



(b)

Fig. 6. Results of the proposed filter (unknown denoising) on different test images, columns 1 through 3 are: Original test image, noisy image, and denoised image a) Test image "Airplane" corrupted with 95% impulse noise b) Test image "Lena" corrupted with 95% impulse noise

[6] P. E. Ng, K. K. Ma, "A Switching Median Filter With Boundary Discriminative Noise Detection for Extremely Corrupted Images," IEEE Trans. Image Process., vol. 15, no. 6, pp.1506-1516, Jun. 2006.

[7] K. Vasanth, T. G. Manjunath, N. Raj "A Decision Based Unsymmetrical Trimmed Modified Winsorized Mean Filter for the Removal of High Density Salt and Pepper Noise in Images and Videos," Procedia Comput. Sci., no. 54, pp. 595-604, 2015.

[8] C. Kandemir, C. Kalyoncu, Ö. Toygar, "A weighted mean filter with spatial-bias elimination for impulse noise removal," Digital Signal Process., vol. 46, pp. 164–174, 2015.

[9] C. T. Lu, Y. Y. Chen, L. L. Wang, , C. F. Chang, "Removal of salt-and-pepper noise in corrupted image using three-values-weighted approach with variable-size window," Pattern Recognition Letters, vol. 80, pp. 188–199, 2016.

[10] H. Hwang and R. A. Haddad, "Adaptive Median Filters: New Algorithms and Results," IEEE Trans. Image Process., vol.4, no. 4, pp. 499-502, 1995.

[11] S. Esakkirajan, T. Veerakumar, A. N. Subramanyam, C.H. PremChand, "Removal of High Dendity Salt and Pepper noise Trhough Modified Decision based Unsymmetrical Trimmed Median Filter," IEEE Signal Process. Lett., vol.18, no.5, pp. 287-290, 2011.

[12] Z. Zhou, "Cognition and Removal of Impulse Noise With Uncertainty," IEEE Trans. Image Process., vol. 21, no. 7, 2012.

[13] C. Kalyoncu1, Ö. Toygar, H. Demirel, "Interpolation-based impulse noise removal," IET Image Process., vol.7, no. 8, pp. 777-785, 2013.

[14] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," IEEE Trans. Image Process., vol. 26, no. 7, pp. 3142-3155, Jul. 2017.

[15] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz, "Loss Functions for Image Restoration with Neural Networks", IEEE Trans. Comput. Imag., vol. 3, no. 1, pp. 47-57, Jan. 2017.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Image net classification with deep convolutional neural networks," In Proc. Advances in Neural Information Processing Systems (NIPS), pp. 1097–1105, 2012.

[17] S. Ioffe, C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," In Proc. International Conference on Machine Learning (ICML), pp. 448–456, Jul. 2015.

[18] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[19] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," IEEE Trans. Image Process., vol. 13, no. 4, pp. 600-612, 2004.

[20] J. Duchi, E. Hazan, and Y. Singer, "Adaptive sub gradient methods for online learning and stochastic optimization," J. Machine Learning Research, vol. 12, no. Jul, pp. 2121–2159, 2011.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," In Proc. IEEE International Conference on Computer Vision, pp. 1026–1034, 2015.